Rapid Identification and Subtyping of Enterobacter cloacae Clinical Isolates Using Peptide Mass Fingerprinting

WANG Yi Qian1,5, XIAO Di1,5, LI Juan1, ZHANG Hui Fang1, FU Bao Qing2, WANG Xiao Ling3, Al Xiao Man4, XIONG Yan Wen1, ZHANG Jian Zhong1,2, and YE Chang Yun1,2*

1. State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Chinese Center for Disease Control and Prevention, Beijing 102206, China; 2. Department of Clinical Laboratory, Daqing Oilfield General Hospital, Daqing 163411, Heilongjiang, China; 3. Department of Clinical Laboratory, Shanxi Province Chinese Medicine Hospital, Taiyuan 030012, Shanxi, China; 4. Department of Medical Laboratory, Beijing Hospital, Beijing 100730, China

Abstract

Objective To establish a domestic database of Enterobacter cloacae (E. cloacae), and improve the identification efficiency using peptide mass fingerprinting.

Methods Peptide mass fingerprinting was used for the identification and subtyping of E. cloacae. Eighty-seven strains, identified based on hsp60 genotyping, were used to construct and evaluate a new reference database.

Results Compared with the original reference database, the identification efficiency and accuracy of the new reference database was greatly improved at the species level. The first super reference database for E. cloacae identification was also constructed and evaluated. Based on the super reference database and the main spectra projection dendrogram, E. cloacae strains were divided into two clades.

Conclusion Peptide mass fingerprinting is a powerful method to identify and subtype E. cloacae, and the use of this method will allow us to obtain more information to understand the heterogeneous organism E. cloacae.

Key words: E. cloacae; Identification; Peptide mass fingerprinting


INTRODUCTION

Enterobacter cloacae (E. cloacae) is a facultatively anaerobic gram-negative bacillus belonging to the Enterobacteriaceae family. This microorganism can be widely found in nature and in saprophytic environments (e.g., soil and sewage); this microbe is also part of the commensal enteric flora of the human gastrointestinal tract. In recent decades, E. cloacae has emerged as a nosocomial pathogen with clinical significance in hospitals[1-6]. In previous work, sequencing of the heat shock protein 60 gene (hsp60) has been helpful for the phylogenetic analysis of Enterobacter. Using hsp60 genotyping, E. cloacae was divided into 12 genetic clusters (cluster I-XII) and

*The study was supported by the Mega Project of Research on the Prevention and Control of HIV/AIDS, Viral Hepatitis Infectious Diseases 2012Z10004-001, 2013Z10004-101 to YE Chang Yun.

*The authors contributed equally to this work.

*Correspondence should be addressed to YE Chang Yun, Professor, PhD, Tel/Fax: 86-10-58900747, E-mail: yechangyun@icdc.cn; ZHANG Jian Zhong, Professor, PhD, Tel/Fax: 86-10-58900754, E-mail: zhangjianzhong@icdc.cn

Biographical notes of the first authors: WANG Yi Qian, female, PhD, majoring in gastrointestinal diseases epidemiology; XIAO Di, female, PhD, majoring in application of proteomics in the diagnosis of infectious diseases.
an unstable sequence cluster (cluster XIII)\(^7\). It is reported that different genetic clusters of \textit{E. cloacae} result in different pathological outcomes\(^8\text{-}13\). However, there is no fast and effective method to identify the genetic clusters of \textit{E. cloacae} from clinical specimens. The BioMerieux biochemical identification system usually confuses \textit{E. cloacae} with \textit{Klebsiella (pneumoniae/oxytoca)} or other species from clinical environments. 16S rDNA sequencing could only identify \textit{E. cloacae} at the species level and is very time consuming and expensive.

Peptide mass fingerprinting (PMF) based on Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) is very useful for the identification of a variety of microorganisms\(^{14\text{-}17}\). Beginning with whole cells, cell lysates, or crude bacterial extracts, the acquired fingerprint spectrum (including signal intensity and mass-to-charge ratio) shows species-specific patterns in cultures grown for a wide range of times using a variety of growth conditions. Through comparison of the acquired spectra with a corresponding reference library, the bacteria can be identified within minutes by analysing the data using various algorithms. With accuracy and automation, high-throughput methods make PMF superior to conventional identification techniques that are based on genome-based identification schemes and biochemical methods\(^{18}\). This study aimed to assess the performance of PMF in the identification and subtyping of \textit{E. cloacae}. PMF data from 86 defined strains were analysed using Biotyper 2.0 software. A new reference database (NRD) and a super reference database (SRD) (including the common characteristics of NRD) were constructed and evaluated. Based on the peptide-mass fingerprints, we created a main spectra projection (MSP) dendrogram and analysed the dendrogram for specific peaks that could be used for the identification and subtyping of \textit{E. cloacae}.

### MATERIALS AND METHODS

#### Bacterial Strains

Eighty-six isolates belonging to 10 genetic clusters of \textit{E. cloacae}, which were isolated from patients in different hospitals, were used in the study (Table 1). Each genetic cluster contained 2 to 16 isolates. A reference strain (ATCC 13047) was also included. All isolates were characterized for clonality using pulsed-field gel electrophoresis and subtyped using \textit{hsp60} genotyping\(^{7,19,20}\). Because no strain from clusters VII, X, and XII was isolated from the clinical samples, our study did not contain those three genetic clusters.

In this study, all strains were isolated from human patients for routine diagnostic purposes. All participants gave written informed consent. This study was approved by the ethics committee of the National Institute for Communicable Disease Control and Prevention, China CDC, according to the medical research regulations of the Ministry of Health, China [Approval No. ICDC-2014003].

#### MALDI-TOF MS Sample Preparation and Data Acquisition

Each strain was selected from brain-heart agar plates. After culturing at 37 °C for 12 h, two samples were prepared from each strain. The extraction method

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No. of strains</th>
<th>Source</th>
<th>Province</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>10</td>
<td>SP (9), Ur (1)</td>
<td>HB, HLJ</td>
<td>2011, 2012, 2013</td>
</tr>
<tr>
<td>III</td>
<td>14</td>
<td>SP (7), Ur (1), SC (5), BL (1)</td>
<td>HB, HLJ</td>
<td>2011, 2012</td>
</tr>
<tr>
<td>IV</td>
<td>5</td>
<td>SP (3), BI (1), OT (1)</td>
<td>HB, HLJ</td>
<td>2011, 2012, 2013</td>
</tr>
<tr>
<td>VI</td>
<td>15</td>
<td>SP (8), Ur (2), SC (3), OT (2)</td>
<td>HB, HLJ, BJ</td>
<td>2011, 2012, 2013</td>
</tr>
<tr>
<td>VIII</td>
<td>16</td>
<td>SP (9), SC (4), BL (1), ST (2)</td>
<td>HB, HLJ</td>
<td>2011, 2012</td>
</tr>
<tr>
<td>IX</td>
<td>5</td>
<td>SP (4), BL (1)</td>
<td>HB, HLJ</td>
<td>2010, 2011, 2012</td>
</tr>
<tr>
<td>XI</td>
<td>2</td>
<td>SP (2)</td>
<td>HB</td>
<td>2011, 2012</td>
</tr>
<tr>
<td>Total</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Sources: SP (sputum), UR (urine), SC (secretion), BI (bile), BL (blood culture), ST (stool), OT (others) Regions: BJ (Beijing), GZ (Guizhou), HB (Hebei), HLJ (Heilongjiang), HN (Henan), QH (Qinghai), SX (Shanxi).
for sample preparation and data acquisition was in accordance with our previously described methods\textsuperscript{[16-17]}. A colony was suspended in 300 μL of molecular-grade water and vortexed, and 900 μL of anhydrous ethanol was added. The samples were vortexed and centrifuged (13,000 × g) for 2 min. The supernatant was discarded and 50 μL of 70% formic acid was added and mixed. Finally, 50 μL of acetonitrile was added, and the solution was carefully mixed. After centrifuging (13,000 × g) for 2 min, the supernatant was the prepared sample. A Microflex LT (Bruker Daltonics) mass spectrometer was used for data acquisition. Escherichia coli strain ATCC 8739 was used for mass calibration and instrument parameter optimization. A Microflex LT instrument was equipped with an N2 laser (λ = 377 nm). The software program used for the data acquisition was FlexControl (version 3.0, Bruker Daltonics). The parameters used were as follows: mass range, 2,000-20,000 Da; ion source 1, 20 kV; ion source 2, 18.5 kV; lens, 8.45 kV; pulsed ion extraction, 330 ns; and laser frequency, 20.0 Hz. Each spectrum was obtained by using 100 shots, and the spectra obtained after 500 shots were superimposed to generate the total spectrum.

Construction and Evaluation of the NRD

In brief, only 17 \textit{E. cloacae} reference strains were present in the original reference database (ORD) of the Biotyper system. Thus, an NRD was constructed using the automated Biotyper (version 2.0) software package\textsuperscript{[21]}. Thirty-four strains belonging to 10 genetic clusters of \textit{E. cloacae} were added to the NRD. The parameters used were as follows: desired mass error for main spectra projection, 200; desired peak frequency minimum, 25%; max, desired peak number for the MSP: 70. For each database entry, 20 individually measured mass spectra were imported into the MSP, which perform normalization, smoothing, baseline correction, and peak picking and generates a list of the most significant peaks. Then, a primary spectrum that contains the average peak mass, the peak intensity and frequency was calculated using the program. Thirty-four reference spectra of 10 genetic clusters were added to the original 17 reference spectra in the ORD; thus, there are 51 reference spectra in the NRD (Table 2). Based on the results of the clustering analysis, 14 strains from clusters III, VI, and VIII were closely related and constituted clade 1, and 20 strains of the other seven clusters were closely related and constituted clade 2. Thus, the strains of clade 1 and clade 2 were compiled as two super reference spectra in the SRD.

Evaluation of the PMF Reference Database

Fifty-three strains were used to evaluate the ORD and NRD. By searching the ORD, thirteen strains with high scores (≥ 2.000) were identified as \textit{E. asburiae}. The other 40 strains were identified as \textit{E. cloacae}, but the score values of eight of these strains were lower than 2.000 (Figure 1).

By searching the NRD, all isolates with high score values (100%, ≥ 2.000) were identified as \textit{E. cloacae}. Only the spectrum for strain CN13EC0078, with a score value of 2.199, matched a previous spectrum; the remaining strains were the first to be matched to the Chinese reference spectrum in the

\textbf{RESULTS}

\textbf{New Reference Database Construction}

Individually measured mass spectra of the \textit{E. cloacae} strains were imported into MSP, which performs normalization, smoothing, baseline correction, and peak picking and generates a list of the most significant peaks. Then, a primary spectrum that contains the average peak mass, the peak intensity and frequency was calculated using the program. Thirty-four reference spectra of 10 genetic clusters were added to the original 17 reference spectra in the ORD; thus, there are 51 reference spectra in the NRD (Table 2). Based on the results of the clustering analysis, 14 strains from clusters III, VI, and VIII were closely related and constituted clade 1, and 20 strains of the other seven clusters were closely related and constituted clade 2. Thus, the strains of clade 1 and clade 2 were compiled as two super reference spectra in the SRD.

\textbf{Evaluation of the PMF Reference Database}

Fifty-three strains were used to evaluate the ORD and NRD. By searching the ORD, thirteen strains with high scores (≥ 2.000) were identified as \textit{E. asburiae}. The other 40 strains were identified as \textit{E. cloacae}, but the score values of eight of these strains were lower than 2.000 (Figure 1).

By searching the NRD, all isolates with high score values (100%, ≥ 2.000) were identified as \textit{E. cloacae}. Only the spectrum for strain CN13EC0078, with a score value of 2.199, matched a previous spectrum; the remaining strains were the first to be matched to the Chinese reference spectrum in

\textbf{MSP Dendrogram and Specific Peak Analysis}

To visualize the relationships among \textit{E. cloacae} strains, an MSP dendrogram was created using the external MATLAB software tool in Biotyper 2.0. The creation of the dendrogram was based on the parameter settings of the standard MSP dendrogram creation method and the general dendrogram settings.

According to the results of the cluster analysis and the super reference spectra, peaks with frequencies greater than 95% were extracted using the Biotyper MSP Peak List Editor (version 2.0.57.0).
Applying peptide mass fingerprinting to *Enterobacter cloacae* 51

While the strains of clusters II, IV, VI, XI, and XIII were correctly identified to the genetic cluster level, misidentifications were identified in strains of clusters I, V, III, VI, and VIII, although these strains had high scores (Figure 1). Strains CN12EC0101 (cluster I) and CN14EC0021 (cluster V) were misidentified as belonging to cluster IV, which suggested that strain CN14EC0017 may be atypical as a reference for cluster IV.

In addition, we constructed and evaluated an SRD for *E. cloacae* identification. Searching the SRD, 30 strains of clusters III, VI, and VIII were identified in clade 1, and twenty-two isolates of seven other genetic clusters were identified in clade 2 (Figure 1).

One strain of cluster III (CN12EC0021) was not reliably identified (score value of 1.642) (Table 2).

### MSP Dendrogram and Specific Peaks

A score-oriented MSP dendrogram was generated using the default settings in Biotyper 2.0 (Figure 2). The 87 strains were categorized into two distinct clades with a distance level of 600. Strains of clusters III, VI, and VIII were categorized as clade 1. Most strains of clade 1 were clustered at the genetic cluster level, with some exceptions. Strains of clade 2 consisted of seven other genetic clusters of *E. cloacae*. The strains clustered at distance levels of 650, 500, 450, 400, 350, 300, and 250 were also

---

**Figure 1.** Comparison of the ORD, NRD, and SRD. The phylogenetic tree based on the *hsp60* gene illustrates the relationship of the 10 clusters of *E. cloacae*. Thirty-four strains were added to the NRD, and another 53 strains were used to evaluate the NRD and SRD. The search results with score values are shown.
### Table 2. Evaluation of Different Reference Databases

<table>
<thead>
<tr>
<th>Cluster</th>
<th>For Construction</th>
<th>For Evaluation</th>
<th>ORD</th>
<th>NRD</th>
<th>SRD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Misidentified</td>
<td>Species Level</td>
<td>Cluster Level</td>
<td>Misidentified</td>
<td>Species Level</td>
</tr>
<tr>
<td>I</td>
<td>3</td>
<td>7</td>
<td>7/100</td>
<td>7/100</td>
<td>7/100</td>
</tr>
<tr>
<td>II</td>
<td>4</td>
<td>3</td>
<td>3/100</td>
<td>3/100</td>
<td>3/100</td>
</tr>
<tr>
<td>III</td>
<td>5</td>
<td>9</td>
<td>9/100</td>
<td>6/68</td>
<td>1/11</td>
</tr>
<tr>
<td>IV</td>
<td>3</td>
<td>2</td>
<td>1/50</td>
<td>2/100</td>
<td>2/100</td>
</tr>
<tr>
<td>V</td>
<td>2</td>
<td>4</td>
<td>4/100</td>
<td>3/75</td>
<td>4/100</td>
</tr>
<tr>
<td>VI</td>
<td>5</td>
<td>10</td>
<td>10/100</td>
<td>7/70</td>
<td>10/100</td>
</tr>
<tr>
<td>VIII</td>
<td>4</td>
<td>12</td>
<td>12/100</td>
<td>9/75</td>
<td>12/100</td>
</tr>
<tr>
<td>IX</td>
<td>3</td>
<td>2</td>
<td>2/100</td>
<td>2/100</td>
<td>2/100</td>
</tr>
<tr>
<td>XI</td>
<td>2</td>
<td>1</td>
<td>1/100</td>
<td>1/100</td>
<td>1/100</td>
</tr>
<tr>
<td>XIII</td>
<td>3</td>
<td>3</td>
<td>2/67</td>
<td>3/100</td>
<td>3/100</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>53</td>
<td>13/25</td>
<td>40/75</td>
<td>53/100</td>
</tr>
</tbody>
</table>

**Note.** ORD, original reference database; NRD, new reference database; SRD, super reference database.

### Figure 2. Cluster analysis of E. cloacae based on the entire protein spectrum (2,000 to 20,000 Da). Strains of clusters III, VI, and VIII were categorized as clade 1, and strains of the seven remaining genetic clusters were clustered as clade 2. Strains that were non-clustered based on their hsp60 genotypes are labelled with asterisks.
classifying to the genetic cluster level (Figure 2). Because cluster XIII was a sequence crowd, the high number of branches suggested that there may be many genotypes in this cluster.

From the MSP of Biotype, fifteen specific peaks with spectra frequencies > 97% were analyzed and extracted from 44 strains of clade 1 (Figure 3A). Eighteen specific peaks with spectra frequencies > 97% were acquired from 43 strains of clade 2 (Figure 3B).

**DISCUSSION**

*Enterobacter cloacae* consists of 13 genetic clusters; the clinical importance and genetic heterogeneity of *E. cloacae* are well known[1,3,6,23-25]. This pathogen usually causes nosocomial wound infections, urinary tract infections, pneumonia, and sepsis in intensive care units. Therefore, a high-throughput diagnostic method would be preferred to the current methods, which are cumbersome. In this study, we constructed and evaluated a new PMF database with hsp60-genotype *E. cloacae* strains; most of the strains were isolated from hospitals in China. The system identified different hsp60 genotypes and clustered *E. cloacae* strains to the genetic cluster level. To the best of our knowledge, this is the first report on the study of *E. cloacae* that was based on a combination of PMF and genetic clustering of *E. cloacae*.

Because of its genetic heterogeneity, *E. cloacae* could be clustered into 13 genetic clusters. In the ORD, there were 17 reference spectra from *E. cloacae* strains; no strain among them was clearly described to the genetic cluster level. In the NRD and SRD, all the added reference strains were typed with the corresponding genetic cluster. Therefore, the NRD and the SRD were more reliable and accurate than the ORD. By including 34 new reference spectra of 10 hsp60-genotype strains, the ability to identify *E. cloacae* strains at the species level improved from

**Figure 3.** Specific peaks of clade 1 and clade 2 from 2,000 to 20,000 Da. (A) Representative pattern of a clade 1 strain (CN11EC0025); (B) Representative pattern of a clade 2 strain (CN12EC0032). The relative intensities of the ions (a.u., arbitrary units) are shown on the Y axis, and the masses (in Da) of the ions are shown on the X axis.
75% to 100%, and at that at the genetic cluster level improved from 0 to 79% (Table 2). In previous reports, some of the genetic clusters have been named: E. asburiae (I), E. kobei (II), E. ludwigi (V), E. hormaechei subsp. oharae (VI), E. hormaechei subsp. hormaechei (VII), E. hormaechei subsp. steigerwaltii (VIII), E. nimipressuralis (X), E. cloacae subsp. cloacae (XI), and E. cloacae subsp. dissolvens (XII); the three remaining clusters are referred to as E. cloacae III, E. cloacae IV and E. cloacae IX\(^{[17]}\). Score values > 2.000 are considered to indicate species level identification for most bacteria\(^{[16]}\); however, this method is poorly suited to E. cloacae identification. Therefore, enriching the capacity of the database and improving the efficiency of recognition are both very important for accurate identification of E. cloacae. Hence, we constructed and evaluated the first version of an SRD for E. cloacae identification. Searching the SRD, 98% of the E. cloacae strains were matched to clade 1 or clade 2. In this study, the identification capabilities of the NRD and SRD for E. cloacae at the species level were roughly the same. However, 42 strains (79%) could be identified at the cluster level by searching the NRD, and 52 strains (98%) could be identified at the clade level by searching the SRD. The normal reference database was constructed using 20 spectra of one strain. The super reference database was constructed using the spectra of all the strains that had common characteristics. In theory, the accuracy of SRD is higher than that of NRD. However, the NRD searching was better for the identification of E. cloacae at the species level, which demonstrated the heterogeneity of E. cloacae.

Based on comparative genomic hybridization, E. cloacae was divided into two clades that are genetically distinct\(^{[19]}\). In our study, the MSP dendrogram clearly shows that clinical isolates can be divided into two distinct clades at the protein level, which is in agreement with the two genetic clades mentioned above. Strains of clusters (III, VI, and VIII) were clustered into clade 1, and the remaining strains were clustered into clade 2. Clade 1 contained the common clinical isolates and was usually isolated from hospital environments with high specific pressures (e.g., antibiotic use). In our previous study, we demonstrated that clusters (III, VI, and VIII) were the most prevalent representatives of E. cloacae in hospital settings (data not shown). Clade 2 contained a relatively large group of species that are associated with the environment or are plant pathogens\(^{[23]}\). Specific peaks were analysed using PMF, which enables discrimination between the two clades. Seven peaks (m/z 3644, 4568, 4753, 5380, 8368, 9137, and 9506) were common in clade 1 and clade 2 (94.8%, 98.9%, 97.9%, 98.9%, 94.8%, and 94.8%, respectively) and considered specific peaks for E. cloacae. Eight peaks (m/z 3133, 4163, 4183, 5149, 5642, 5669, 8294, and 8327) were specific for the identification of clade 1. Eleven peaks (m/z 3125, 3620, 4361, 4443, 5404, 6255, 6282, 6330, 7244, 7812, and 8893) were specific for the identification of clade 2. Therefore, further study of the different protein sequences between the two clades may help explain why clusters III, VI, and VIII are prevalent in hospitals. Strains from clade 2 were clustered at different distance levels, thus, the specific peaks corresponding to different genetic clusters warrant further study.

In recent years, many studies have used biochemical and molecular methods to focus on rapid identification and subtyping\(^{[19,26]}\). Comparative genomic hybridization (CGH), multi-locus sequence analysis (MLSA), multi-locus sequence typing (MLST), and hsp60 and rpoB genotyping are better methods for the identification and subtyping of E. cloacae species; however, these methods are expensive and labour intensive. Peptide mass fingerprinting is an emerging technique for the rapid identification of microorganisms\(^{[27-29]}\). Our study shows that PMF is a powerful method for studying E. cloacae. Using PMF, the proteins could be tested immediately after the strains were cultured and isolated. The entire process, from protein extraction to reference database searching (NRD and SRD), could be performed within 10 min. Based on the PMF data, the MSP dendrogram can be used for identification and subtyping of E. cloacae. This method is high throughput and low cost. Therefore, we believe that the PMF method is equivalent or superior to traditional diagnostic methods for E. cloacae. Thus, the PMF method is a novel and powerful method to study genetic cluster assignments within E. cloacae and is able to detect specific genetic clusters that threaten hospital populations. In combination with other genotyping methods, this method could be used to analyse the relationships among the genetic clusters of E. cloacae, greatly improving identification and molecular subtyping.

ACKNOWLEDGEMENTS

For excellent technical assistance, we thank ZHAO Ai Lan, WANG Yan, JIN Dong, YUAN Xue Jiao, YU Bo, MENG Shuang, and BAI Xiang Ning. In
Applying peptide mass fingerprinting to Enterobacter cloacae

addition, we thank LIU Kai and LI Pei Jing for material support. The authors are grateful to WANG Yi for careful proofreading of the manuscript.

AUTHOR CONTRIBUTIONS

WANG Yi Qian and YE Chang Yun designed the project and wrote the paper. WANG Yi Qian, XIAO Di, XIONG Yan Wen, and ZHANG Jian Zhong carried out the experimental work. LI Juan, ZHANG Hui Fang, FU Bao Qing, WANG Xiao Ling, and AI Xiao Man isolated Enterobacter cloacae strains from samples. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interests.

Received: August 15, 2017;
Accepted: December 13, 2017

REFERENCES


